Chapter 2

Measures of Central Tendency

Learning Objectives

- Understand how frequency tables are used in statistical analysis.
- Explain the conventions for building distributions.
- Understand the mode, median, and mean as measures of central tendency.
- Identify the proper measure of central tendency to use for each level of measurement.
- Explain how to calculate the mode, median, and mean.

Key Terms

bivariate	median
central tendency	mode
frequency distribution	multivariate
mean	univariate

Now we begin statistical analysis. Statistical analysis may be broken down into three broad categories:

- Univariate analyses
- Bivariate analyses
- Multivariate analyses

These divisions are fairly straightforward. *Univariate* analyses deal with one variable at a time. *Bivariate* analyses compare two variables to each other to see how they differ or if they are the same. In statistical analysis this could take the form of association or correlation. *Multivariate* analyses compare three or more variables to one another, or compare more than two variables (called *independent variables*) in how they influence a variable we are interested in knowing something about (called a *dependent variable*).

The focus of this chapter and of Chapter 3 is on univariate analysis. Univariate descriptive statistics are used to describe and interpret the meaning of a distribution. These statistics tell a lot about a data set by revealing a few critical characteristics. You should always undertake this process before doing any bivariate or multivariate analyses; but you can do sound research just using the analyses we will address in these chapters. Univariate descriptive statistics make compact characterizations of distributions in terms of three properties of the data: *central tendency*, which translates to the average, middle point, or most common value of the distribution; *dispersion*, which relates to how spread out the values are around the central measure; and the *form* of the distribution, which relates to what the distribution would look like if it were displayed graphically.

2-1 Frequency Distributions

Most statistical analyses are displayed through tables (and sometimes charts). Throughout these chapters, you will see different kinds of tables and graphs displaying various types of data. At the most basic level, data is displayed through a frequency distribution (which often contains univariate statistics displayed with the data). Frequency distributions are helpful in summarizing data (scores) because they provide a visualization of how scores are spread over categories. The most common type of distribution for current statistical packages is a combination distribution. This is the type of distribution shown in **Table 2-1**. In SPSS a combination distribution includes percentage, valid percentage, and cumulative percentage. A lot can be learned from these frequency distributions, but we want to focus on the univariate descriptive statistics that usually are a part of these printouts.

Conventions for Building Distributions

Within statistical analysis, there are conventions and standards for presenting tables and distributions. You will soon realize that many people do not follow these conventions; however, if you adhere to these guidelines when building your own tables and distributions, they will be clearer and the results will be easier to interpret.

First, the title of the table or distribution should clearly explain the contents. It should contain the variables included in the table and any variable label or explanation

EDUCATION What is your highest level of education?					
Value Label	**Value**	**Frequency**	**Percent**	**Valid Percent**	**Cumulative Percent**
Less than High School	1	16	4.6	4.8	4.8
GED	2	59	17.0	17.6	22.3
High School Graduate	3	8	2.3	2.4	24.7
Some College	4	117	33.7	34.8	59.5
College Graduate	5	72	20.7	21.4	81.0
Postgraduate	6	64	18.4	19.0	100.0
Total		336	96.8	100.0	
System	Missing	11	3.2		
	Total	347	100.0		

Table 2-1 Combination Table for Education from 1993 Little Rock Community Policing Survey

EDUCATION What is your highest level of education?					
Value Label	**Value**	**Frequency**	**Percent**	**Valid %**	**Cum %**
Less than High School	1	16	4.6	4.8	4.8
GED	2	59	17.0	17.6	22.3
High School Graduate	3	8	2.3	2.4	24.7
Some College	4	117	33.7	34.8	59.5
College Graduate	5	72	20.7	21.4	81.0
Postgraduate	6	64	18.4	19.0	100.0
System	Missing	11	3.2		
		347	100.0		

Table 2-2 Constructed in Excel

of the variable that makes it clear what data the distribution contains. In Table 2-1, the variable is *education*, and the explanation is the wording from the survey from which the variable was drawn: "What is your highest level of education?" If the origin of the data is not expressed clearly in the discussion, the title should also contain this information. This may include the year the data was collected, the geographic region of collection, and any specific information about how the data was gathered.

Possibly more important than the title is proper labeling of the columns and categories. The categories of the data should be clearly identified in the left column, called the *stub* of the table or distribution. These categories should be labeled even if the values are included. In Table 2-1, simply including the values 1 to 6 does not help the reader interpret the table unless he or she has the categories (less than high school, GED, high school, etc.) that correspond to the values. The words that represent the numbers in the stub are called the *value labels*. Each column should also be clearly labeled with its heading.

All relevant columns should be totaled. These totals are called *marginals*. This is important information for summary analysis, and it prevents the reader from having to make independent calculations. It is also important to provide the reader with information on how the table or distribution was developed. In Table 2-1, for example, the frequency and percent totals are based on 347 responses. Valid percent and cumulative percent, however, are based on the total minus the missing responses (11).

Finally, footnotes should be added to tables and distributions when more information is needed, when clarification is necessary, or when there is something unusual about the table or distribution. If not included in the title, the source of the information should be presented in a footnote. This is particularly important when the data in the table or distribution is not collected by the presenter.

Box 2-1 How Do You Do That?

Throughout this text, we will present boxes that show you how to obtain the results or output discussed in the chapter. Often, both SPSS and Excel methods will be demonstrated. For more advanced analyses, only SPSS will be shown. This box shows you how to obtain *frequency distributions*.

SPSS

1. Open a data set.
   a. Start SPSS.
   b. Select File, then Open, then Data.
   c. Select the file you want to open, then select Open.
2. Once the data is visible, select Analyze, then Descriptive Statistics, then Frequencies.
3. Make sure that Display Frequency Tables is checked.
4. Select the variables you wish to include in your distribution, and press the ▶ between the two windows.

5. For now, do not worry about the option boxes at the bottom of the window, and just press OK.
6. An output window should appear that contains a distribution similar in format to Table 2-1.

Excel

Excel does not build tables for you. To complete a table in Excel, follow the instructions for building a table described above, filling in the proper cells with the values and/or data.

1. Start in the second row and fill in the column headings. Block the cells, select the Bold button, and then then select the Center button to make the text boldface and to center it.
2. Block the cells above the headings, right-click the mouse and select Format Cells, select the Alignment tab, and check the box marked Merge Cells. Type in the heading of the table and select the Center button to center the text.
3. Type in the value labels. Note that you may have to expand the cell width to get all of the text in the cell.
4. Type in the data values and the frequencies. You might place the word *Missing* in the cell identifying missing data so you know immediately what that is.
5. Now Excel will do calculations for you. To total the frequencies, click on the Σ in the Home toolbar. Block the cells you want to add and press Return.
6. To calculate the Percent, click on the Σ in the Home toolbar, then select the first frequency (16), type in a / in the formula bar next to the cell value (C3), then type in the frequency total (347), then press Return. You will need to right-click the cell, select Format Cells, select the Number tab, select Percentage from the list, then make sure it is 1 decimal place. To calculate the percentages for the other cells, simply block the cell and then drag down to copy. To make sure you have 100%, simply add the column as described in step 5.
7. Calculating the Valid Percent is done in much the same way as calculating the Percent. The difference here is using $347 - 11$ to divide each cell by the total minus the missing values.
8. To calculate the Cumulative Percent, copy and paste the value in the first cell in the Valid Percent column (4.8%). Click on the Σ in the Home toolbar, then select the value above the empty cell (4.8%), then type a + into the formula bar, then select the cell next to the empty cell, then press Return. This will give the next value in Cumulative Percent. You can fill in the rest of the cells by copying and pasting this cell.

The result of this work will look like **Table 2-2**. Save this table once you have created it. It is a simple matter to replace these values with others to quickly make additional frequency tables in Excel.

2-2 Measures of Central Tendency

Measures of central tendency examine where the distribution's central—or most typical—value is located. There are three common measures of central tendency, one for each level of measurement. These are the *mode* for nominal level data, the *median* for ordinal level data, and the *mean* for interval and ratio level data.

Mode

At the lowest level of sophistication is the *mode* (symbolized by Mo). The mode is used primarily for nominal data to identify the category with the greatest number of cases. The mode is the most frequently occurring value, or case, in a distribution. It is the tallest column on a histogram or the peak on a polygon or line chart. The mode has the advantage of being spotted easily in a distribution, and it is often used as a first indicator of the central tendency of a distribution.

The mode is the only measure of central tendency appropriate for nominal variables because it is simply a count of the values. Unlike other measures of central tendency, the mode explains nothing about the ordering of variables or variation within the variables. In fact, the mode ignores information about ordering and interval size even if it is available. So it is generally not advised to use the mode for ordinal or interval level data (unless it is used in addition to the median or mean) because too much information is lost.

There is no formula or calculation for the mode. The procedure is to count the scores and determine the most frequently occurring value. Consider the data set in **Figure 2-1**, which show the number of prisoners who escaped from 15 prisons over a 10-year period. Two prisons had 7 escapes, one had 6 escapes, three had 5, two had

7	5	4	3	2
7	5	4	3	1
6	5	3	3	1

Figure 2-1 Number of Escapes from 15 Prisons, 2000–2010

4, four had 3, one had 2, and two had 1 escape. The mode in this data set would be 3 escapes, since there are more 3s than any other value.

For data that is already in a frequency table, determining the mode is even easier because the numbers are already counted. In Table 2-1, you can easily see that Some College (value 4) has the greatest number of people (117). This means that 4 or Some College is the mode.

One caution when discussing the mode. The mode is *not* the frequency of the number that occurs most often but is rather the category (or class) itself. It is tempting to state that the mode in Table 2-1 is 117 because that is the frequency that is highest. This is not the mode, however; the mode is the *category* of the value that has the highest frequency—in this case, 4, or Some College.

A distribution is not confined to having only one mode. There are often situations in which a distribution will have several categories that have the same or similar frequencies. In these cases, the distribution can be said to be *bimodal* or even *multimodal*. It is also possible for a distribution to have no mode if the frequencies are the same for each category.

Box 2-2 Obtaining Univariate Descriptive Measures in SPSS

You can obtain all of the univariate descriptive statistics within the same procedure in SPSS, and it is just an extension of the same procedure used above to obtain a frequency distribution. Follow these steps:

1. Open a data set such as one provided at go.jblearning.com/Walker.
   a. Start SPSS.
   b. Select File, then Open, then Data.
   c. Select the file you want to open, then select Open.
2. Select Analyze, then Descriptive Statistics, then Frequencies.
3. Make sure that Display Frequency Tables is checked.
4. Select the variables you wish to include in your distribution and press the > between the two windows.
5. Select the Statistics button at the bottom of the window.
6. Check the boxes of any of the univariate measures you want to include.
7. Select Continue, then OK.
8. An output window should appear containing a distribution similar in format to **Table 2-3**.

Median

If the data is at least ordinal level, the median (symbolized by Me) may be a better choice for examining the central tendency of the distribution. The *median* is the point of the 50th percentile of the distribution. This means that the median is the exact midpoint of a distribution, or the value that cuts the distribution into two equal parts.

For the simple distribution 1, 2, 3, the median is 2 because it cuts this distribution in half. Note that 2 is not the most frequently occurring value or the product of some formula, but is simply the value in the middle. The median will always be the middle value, but sometimes it will be necessary to resort to math to determine the exact middle value.

EDUCATION What is your highest level of education?

Value Label	Value	Frequency	Percent	Valid Percent	Cumulative Percent
Less than High School	1	16	4.6	4.8	4.8
GED	2	59	17.0	17.6	22.3
High School Graduate	3	8	2.3	2.4	24.7
Some College	4	117	33.7	34.8	59.5
College Graduate	5	72	20.7	21.4	81.0
Postgraduate	6	64	18.4	19.0	100.0
Total		336	96.8	100.0	
System	Missing	11	3.2		
	Total	347	100.0		

N	Valid	336
	Missing	11
Mean		4.08
Median		4.00
Mode		4
Std. Deviation		1.460
Variance		2.131
Skewness		−0.477
Std. Error of Skewness		0.133
Kurtosis		−0.705
Std. Error of Kurtosis		0.265
Range		5

Table 2-3 SPSS Output of Univariate Descriptive Statistics

The median is used with ordinal level data because it does not imply distance between intervals, only direction—above the median or below it. Recall from Chapter 1 that the nature of ordinal level data is that you can determine which category is greater than or less than another category; but there are not equal intervals, so there is no way to determine how much greater or lesser the category is. The median also works on this principle, determining the midpoint of a distribution such that a category can be said to be less than or greater than the median, but there is no way to tell by how much. For example, take the following two distributions:

1, 2, 3, 3, 4, 4, 5   1, 1, 1, 3, 10, 50, 100

Each has the same number of values, 7, although each has very different numbers. In this case, the modes are different: 3 and 4 in the first; 1 in the second. Also, as discussed below, the means are different: 3.14 in the first, 23.71 in the second. The median for both of these distributions, however, is 3, the middle value in the distribution. In both distributions, there are three values below the median and three values above the median.

Calculation of the median is relatively simple. All that is needed is N for the distribution. If N is not given, simply count the number of scores (remember—do *not* add the scores, *count* them). Then the value of N is placed in the formula $(N + 1)/2$. Once you have this value, if the scores are not arranged in order, you should do so. Then simply count up the data until you reach the value obtained in the formula (the 12th value in **Figure 2-2**). This is the median. In this case, if we count to the 12th score, we get a score of 3. An example of how to make the same calculations in Excel is shown in Box 2-3.[1]

7	5	4	2	2
6	5	3	2	1
6	5	3	2	1
6	5	3	2	
6	4	2	2	

$$\frac{N+1}{2} = \frac{23+1}{2} = \frac{24}{2} = 12$$

Figure 2-2 Calculating the Median

Box 2-3 Using Excel to Calculate the Median

1. If the data is not sorted, do so by clicking on the column heading for the data you want to sort, clicking on Data, then clicking on the button to Sort Descending.
2. Since there are 23 rows, this is *N*.
   a. N + 1 = 24.
   b. 24 / 2 = 12.
3. Count up (or down) in the data until you reach the 12th number. This is 3, which is the median.

There are several issues to note about the median. These should be considered any time you use the median for analysis. First, you must be careful when you calculate the median because there are two different numbers you must deal with. To begin, you get a number from the formula. This is not the median; it is simply the number of values to count up in the distribution to find the median. The median is the score that contains the number from the formula. Also, if there is more than one of the same score in the median class (there are three 3s in Figure 2-2), the median is still that score even though it occurs more than once. Finally, unlike the mode, the median does not have to be a value that is in the distribution. For an odd number of scores (such as in Figure 2-2), the median will be one of the scores because it is the point that cuts the distribution in half. If there are an even number of scores, however, the median will fall in between two of the scores. For example, in the distribution

$$3, 4, 5, 6, 7, 8, 9, 10$$

the formula $(N + 1)/2$ gives a value of 4.5. This puts the median between 6 and 7. When this occurs, the median is a value that is halfway between the two scores. In this case, the median is 6.5. This is even more complicated if the two numbers do not have an interval of 1. For example, in the distribution 5, 6, 8, 10, 11, 12, the number from the formula puts the median between the score of 8 and the score of 10; therefore the median is 9.

Mean

The most popular measure of central tendency is the *mean*. The mean is primarily used for interval and ratio level data. Because it assumes equality of intervals, the mean is generally not used with nominal or ordinal level data. The mean is very important to statistical analysis because it is the basis (along with the variance) of many of the formulas for higher-order statistical procedures.

The symbolic notation for the mean is different than others that have been used to this point. The mean is symbolized either by μ or $\overline{X}$, depending on whether the data is a population or sample estimate (this distinction will be discussed when we get to inferential statistics). It is interesting that descriptive statistics deals with a population, but it has become convention that the mean most commonly used in descriptive statistics is actually the symbol for the sample mean ($\overline{X}$). Since most texts use this notation for the mean in descriptive analyses, it will also be used here, even though the more proper notation is the population mean (μ).

The mean is simply the average of the values in the distribution. To obtain the mean, add the scores in a distribution and divide by *N* (it is just like calculating an average). In statistical terms, the mean is calculated as

$$\overline{X} = \frac{\Sigma fX}{N}$$

If we grouped the data from Figure 2-2, it would look like that in Table 2-4. Here, all we did is to take the numbers and add a column called *f*. This column is the number of times that value occurs in our data set. In this table, we also added a column called *fX*, where *fX* is calculated by multiplying *X* by the frequency *f* for each value. If we add the numbers in the *fX* column, we get 84. If we add the numbers in the *f* column, we get 23, which is the same as *N*. So the sum of *f* is the same as *N*. Now we have all of the numbers we need for the formula. Plugging the numbers into the formula, we get 84 ÷ 23, or 3.65.

The mean has some advantages over other measures of central tendency. From a practical standpoint, the mean is preferred because it is standardized, allowing for it to be compared across distributions. This is beneficial when you are comparing similar

X	*f*	*fX*
7	1	7
6	4	24
5	4	20
4	2	8
3	3	9
2	7	14
1	2	2
N	23	
ΣfX		84

Table 2-4 Calculating the Mean

data from different sources. For example, we could compare the mean number of calls for service among several police agencies because we can directly compare their mean values and tell which agency had more calls for service on average.

The greatest problem with the mean is that it is highly affected by extreme scores in the distribution. In the following data

1, 1, 3, 3, 4, 4, 100

the median is 3, which is reasonable given this data. The mean, however, is 16.57. This number is not even close to most of the numbers in the data set; but because one of the numbers is 100, it draws the mean away from the other numbers. That is why the median is used in cases where data is skewed, as discussed in Chapter 3.

2-3 Selecting the Most Appropriate Measure of Central Tendency

As with levels of measurement, the most important part of measures of central tendency is the selection of the proper one for the data. Selection of the most appropriate measure of central tendency depends on several factors, including the level of data, the nature of the data, and the purpose of the summarization.

The level of data has a substantial influence on which measure of central tendency should be used. As discussed above, each measure is most appropriate for a particular level of data. The mode is most appropriate for nominal level data, and its use with ordinal and interval level data results in a loss of power in terms of the information that could be gained from the data. The median is most appropriate with ordinal level data; and although it can be used with interval level data (especially skewed distributions), it should not be used with nominal level data because the rankings assumed in the median cannot be achieved with nominal level data. Finally, the mean should only be used with interval and ratio level data because it assumes equal intervals of the data that cannot be achieved by nominal and ordinal level data.

Selection of the most appropriate measure of central tendency is also sometimes based on the nature of the distribution. As will be discussed more fully in Chapter 3, if a distribution is highly skewed, the median should be used rather than the mean for interval and ratio level data.

The final criterion for choosing a measure of central tendency is the purpose of summarization, typically in terms of what you are trying to predict. Imagine that you are asked to state one measure that will best capture the nature of the distribution. How do you go about that? To put it another way, you might bet $100 that you could guess a number drawn at random from a distribution. Which one do you choose? One way to address these questions is to find the score at the "heart" of the distribution—the most common score, the one that cut the distribution in half, the average score. That is the goal and role of measures of central tendency.

If you know all the values in the distribution, you can easily and quickly calculate the mode. If you are interested in predicting an exact value, you should probably use the mode since it has the highest probability of occurring in any given distribution. Both the median and the mean have the possibility to return values that are not in the distribution; so if you must guess and be absolutely right on the number, use the mode.

If, on the other hand, you want to maximize your prediction by getting closest to the number over several tries, thereby minimizing your error, then the median may be a better choice. Here, whether you chose too high or too low is irrelevant; what is important is the size of the error. In a popular game show, contestants are given $7 and required to guess the exact numbers included in the price of a car. For each number they are off, they lose $1. If they have money left over after making all of the guesses, they win the car; if they run out of money, they lose. Probability of response plays a big part in the first two or three numbers (you would not want to guess 9 for the first number, for example). If you are at the fourth or fifth number, however, and still have money left, you may want to choose the median value (probably a 5) to minimize the error (loss of dollars). The median is both good and bad in this instance. It is good in that, since it is the middle value, it will produce the smallest absolute error from any value you choose. There is, however, an equal chance of its being high or low on any given value chosen.

Finally, if you have the opportunity to average your misses over several guesses and the signs do matter (high guesses can offset low guesses), then the mean is the best choice. The mean is good in that if you don't know a value, it is often best to choose the average.

This chapter addressed constructing and understanding frequency distributions and measures of central tendency. These skills will be necessary in Chapter 3 as the next two univariate descriptive statistics are discussed.

2-4 Equations Used in This Chapter

Mean:

$$\overline{X} = \frac{\sum fx}{N}$$

Median:

$$\text{Me} = \frac{(N + 1)}{2}$$

2-5 Note

1. Note that it is possible to calculate the median and other statistics directly in Excel. This approach has received some criticism from the American Statistical Association and others. As such, this text will use Excel to aid in calculations, but will not utilize its formulas or pivot table options. These procedures may be taught in class, however.

2-6 For Further Reading

Edwards, W. J., White, N., Bennett, I., & Pezzella, F. Who has come out of the pipeline? African Americans in criminology and criminal justice. *Journal of Criminal Justice Education*, 1998, *9*(2), 249–266.

Galton, F. *inquiries into human faculty and its development.* London: MacMillan, 1883.

Pearson, K. Classification of asymmetrical frequency curves in general. Types actually occurring. *Philosophical Transactions of the Royal Society of London*, series A, vol. 186. London: Cambridge University Press, 1895.

2-7 Exercises

The exercises for this chapter and for Chapter 3 will use the same examples. This will allow you to work through problems using all three types of univariate descriptive statistics.

1. For the set of data below, calculate the
   a. Mode
   b. Median
   c. Mean

   6, 7, 8, 10, 10, 10, 12, 14

2. For the set of data below, calculate the
   a. Mode
   b. Median
   c. Mean

   7, 4, 2, 3, 4, 5, 8, 1, 9, 4

3. For the set of data below, calculate the
   a. Mode
   b. Median
   c. Mean

Value (*X*)	Frequency (*f*)
90	6
80	8
70	4
60	3
50	2

4. For the set of data below, calculate the
   a. Mode
   b. Median
   c. Mean

Value (*X*)	Frequency (*f*)
90	5
80	7
70	9
60	4

5. Using the following frequency tables, discuss the three measures of central tendency.

HOME What type of house do you live in?					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid		1	.3	.3	.3
House	1	279	81.3	82.1	82.4
Duplex	2	3	0.9	0.9	83.2
Trailer	3	34	9.9	10.0	93.2
Apartment	4	21	6.1	6.2	99.4
Other	5	2	0.6	0.6	100.0
	Total	340	99.1	100.0	
Missing	System	3	0.9		
Total		343	100.0		

N	Valid	340
	Missing	3
Mean		1.41
Std. Error of Mean		0.051
Median		1.00
Mode		1
Std. Deviation		0.945
Variance		0.892
Skewness		2.001
Std. Error of Skewness		0.132
Kurtosis		2.613
Std. Error of Kurtosis		.264
Range		5

ARREST How many times have you been arrested?					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	0	243	70.8	86.2	86.2
	1	23	6.7	8.2	94.3
	2	10	2.9	3.5	97.9
	3	3	0.9	1.1	98.9
	5	2	0.6	0.7	99.6
	24	1	0.3	0.4	100.0
	Total	282	82.2	100.0	
Missing	System	61	17.8		
Total		343	100.0		

N	Valid	282
	Missing	61
Mean		0.30
Std. Error of Mean		0.093
Median		0.00
Mode		0
Std. Deviation		1.567
Variance		2.455
Skewness		12.692
Std. Error of Skewness		0.145
Kurtosis		187.898
Std. Error of Kurtosis		0.289
Range		24

TENURE How long have you lived at your current address (months)?

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	14	4.1	4.3	4.3
	2	6	1.7	1.8	6.1
	3	4	1.2	1.2	7.3
	4	4	1.2	1.2	8.6
	5	6	1.7	1.8	10.4
	6	6	1.7	1.8	12.2
	7	1	0.3	0.3	12.5
	8	3	0.9	0.9	13.5
	9	2	0.6	0.6	14.1
	10	1	0.3	0.3	14.4
	11	1	0.3	0.3	14.7
	12	11	3.2	3.4	18.0
	14	1	0.3	0.3	18.3
	18	5	1.5	1.5	19.9
	21	1	0.3	0.3	20.2
	24	30	8.7	9.2	29.4
	30	1	0.3	0.3	29.7

TENURE How long have you lived at your current address (months)?					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	31	1	0.3	0.3	30.0
	32	1	0.3	0.3	30.3
	36	22	6.4	6.7	37.0
	42	1	0.3	0.3	37.3
	48	12	3.5	3.7	41.0
	60	24	7.0	7.3	48.3
	72	14	4.1	4.3	52.6
	76	1	0.3	0.3	52.9
	84	8	2.3	2.4	55.4
	96	18	5.2	5.5	60.9
	108	4	1.2	1.2	62.1
	120	9	2.6	2.8	64.8
	132	11	3.2	3.4	68.2
	144	21	6.1	6.4	74.6
	156	13	3.8	4.0	78.6
	168	11	3.2	3.4	82.0
	170	5	1.5	1.5	83.5
	180	7	2.0	2.1	85.6
	182	2	0.6	0.6	86.2
	186	1	0.3	0.3	86.5
	192	14	4.1	4.3	90.8
	198	1	0.3	0.3	91.1
	204	24	7.0	7.3	98.5
	216	3	0.9	0.9	99.4
	240	2	0.6	0.6	100.0
	Total	327	95.3	100.0	
Missing	System	16	4.7		
Total		343	100.0		

N	Valid	327
	Missing	16
Mean		88.77
Std. Error of Mean		3.880
Median		72.00
Mode		24
Std. Deviation		70.164
Variance		4,923.055
Skewness		0.365
Std. Error of Skewness		0.135
Kurtosis		−1.284
Std. Error of Kurtosis		0.269
Range		239

SIBS How many brothers and sisters do you have?					
		Frequency	**Percent**	**Valid Percent**	**Cumulative Percent**
Valid	0	39	11.4	11.5	11.5
	1	137	39.9	40.5	52.1
	2	79	23.0	23.4	75.4
	3	39	11.4	11.5	87.0
	4	17	5.0	5.0	92.0
	5	13	3.8	3.8	95.9
	6	6	1.7	1.8	97.6
	7	4	1.2	1.2	98.8
	9	1	0.3	0.3	99.1
	10	1	0.3	0.3	99.4
	12	1	0.3	0.3	99.7
	15	1	0.3	0.3	100.0
	Total	338	98.5	100.0	
Missing	System	5	1.5		
Total		343	100.0		

N	Valid	338
	Missing	5
Mean		1.94
Std. Error of Mean		0.098
Median		1.00
Mode		1
Std. Deviation		1.801
Variance		3.245
Skewness		2.664
Std. Error of Skewness		0.133
Kurtosis		12.027
Std. Error of Kurtosis		0.265
Range		15